- Home
- Calendar
- Search
- Contacts

Tracking SARS-CoV-2 Spike mutations

A SARS-CoV-2 variant carrying the Spike protein amino acid change D614G has become the most prevalent form in the global pandemic.

LANL'S BROAD EFFORT TO FIGHT COVID-19

Fact sheet

TRACKING SARS-COV-2 SPIKE MUTATIONS

Read the paper

SARS-COV-2 PAPER COMMENTARY

Making sense of the mutation

CONTACT

Nancy Ambrosiano
(505) 699-1149
Email

Charles Poling

For more information, jump to:

SARS-CoV-2 mutation: You've got questions, we've got answers

This is a synthesis of questions Los Alamos National Laboratory researchers and their colleagues have received both from scientists and from the press regarding the paper "Tracking Changes in SARS-Cov-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 virus, Korber et al., 2020, Cell 182, 1-16.

In this paper, the Los Alamos team and their colleagues provided evidence that particular SARS-CoV-2 mutation was associated with increased viral transmission and the spread of COVID-19, was more infectious in cell culture, and was associated with higher levels viral genetic material in the upper respiratory tract of infected individuals. The variant in question, D614G, makes a small but effective change in the virus's 'Spike' protein, which the virus uses to enter human cells.

Questions

(505) 257-8006
Email

Laura Mullane
(505) 412-7733
Email

1. What does D614G mean?

D614 means the original form; the mutant form is referred to as *D614G*, or just *G614*.The coronavirus that causes COVID-19 uses its Spike protein to infect human cells. D614G refers to an amino acid mutation in this protein that has become increasingly common in SARS-CoV-2 viruses from around the world. The Spike protein (S) is a string of 1,273 amino-acids; in the original form from Wuhan the 614th of these amino acids has the chemical symbol "D" (aspartic acid), while in the mutated form, the 614th amino acid is abbreviated "G" (glycine); so S D614G is short for "having a Spike protein with aspartic acid at position 614 mutated to glycine."

2. Is the D614G mutation recurring in many individuals, or is it being transmitted from person to person?

The vast majority of the time it is being transmitted person to person, and not arising independently as a new mutation. The D614G mutation is being carried along as a part of a clade called the "G clade" by GISAID that is named for this mutation. A "clade" is a lineage in a phylogenetic tree with a shared ancestral state. A phylogenetic tree can be thought of like a family tree, and a shared ancestor like a grandmother who is a shared ancestor of all of her grandchildren.

The G clade differed from the original Wuhan form by 4 mutations. G614 is almost always found linked to the other 3 mutations (>99.99% of the time, Korber et al. 2020, Fig. S5).

There are a small number of GISAID sequences where the 4 base haplotype is disrupted and not all 4 bases are present. While some of these may have

been spontaneous dead-end mutations, early examples may have been ancestral to the G clade lineage. Among the later samples, at least some of these result from recombination, and again, not as *de novo* mutation (Korber et al. 2020, bioRxiv).

3. What was the evidence indicating that the D614G mutation is more transmissible than the original form? What is the current evidence?

The genetic evidence:

The Wuhan form of the virus rapidly spread throughout the globe in early 2020. In local geographic populations where both the G clade and the original Wuhan form co-circulated, the G form repeatedly showed rapid and significant increases in relative frequency. This pattern was consistently repeated at virtually every geo/political level: country, state, county, and city, with only very rare exceptions. At the time of our bioRxiv submission (Korber et al. 2020, bioRxiv) we had found several dozen cases where the G form was increasing, and only a single exception; we made a public website (cov.lanl.gov) to enable other people to track the D614G frequency relative to the original Wuhan form in any area in the world based on GISAID data. GISAID is the main database and global repository of SARS-CoV-2 genetic sequences, and teams from all over the world provide the data; currently it houses 80,000 SARS-CoV-2 sequences (gisaid.org, Aug. 5, 2020).

By the time of our *Cell* publication, we had developed two systematic strategies to explore and analyze all geographic regions in GISAID with enough data (enough sequences sampled over enough time) to look for frequency changes (Korber et al. 2020, Fig. 1B and 3). We found the G clade had significantly increased in frequency in 47/50 such geographic regions (Korber et al. 2020).

As of Aug., 2020, the G clade has been found to significantly increase in frequency over time in 121 geographic locations, while the original form increased in only 3 locations (isotonic regression, update of Fig. 3 in the Cell paper). The 121 geographic regions include 32 countries, 58 states and provinces, 31 cities and counties.

Founder effects and random events might indeed cause an observed frequency shift from one form to the other in a given region. This is an important point, and it is often brought up in the debate about whether our results were indicative of a change in transmissibility. A founder event might be a super-spreader event (like a crowded concert) or a new introduction into new regions that happened to take off. A typical formulation was, “But these are just new introductions seeded from New York.”

But random events are by definition just that, random. If the two forms are equally likely to propagate, *shifts in frequency would be (more or less) evenly balanced: sometimes D-to-G, sometimes G-to-D. You would not expect the shift to almost always go in the same direction.* But that is exactly what happened: The shifts almost always went in one direction, towards higher G clade frequencies. The multiple independent repetitions of the pattern provided compelling evidence of positive selection, and we have emphasized this in all of our writing about and talks about the D614G mutation. This was the central point of our paper.

The frequency shifts towards the G clade throughout March were happening in areas everywhere, including many areas where the G clade was introduced into extremely well-established local D clade epidemics. This is very evident when you look at the plots of the in frequencies over time. Australia, Japan, Hong Kong, Spain, Thailand, and the UK are a few examples at a national level, and there many more examples at more local levels (Korber et al. 2020, cov.lanl.gov).

Here is the breakdown for two examples, counties from Washington State (also see Korber et al. 2020, Fig. Sup. 2): ((Confirmed COVID19 case count data from: COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University)

Snohomish County, Washington State, USA

Stay-at-home orders were given on March 24.

There were 614 confirmed cases by March 24, (COVID-19 Data Repository Johns Hopkins University), and the number of infections per reported case in western Washington in this time frame was estimated to be 11.2 to 1. (Havers et al, 2020). Thus, there were likely to have been ~6,900 COVID infections in King County by March 24.

33 of the cases through March 24 were sequenced, and 100% of were the D form at position 614 (GISAID data).

This was obviously a well-established D clade epidemic. But by early April almost all of the infections sampled were G clade.

Fig 1. A plot of cumulative counts of D614 and G614 sequences by day in Snohomish county. Orange represents the original form, blue the form with the G614 mutation. 2020.

King County, Washington state, USA

Stay at home orders were given on March 24.

There were 1,170 confirmed cases by March 24, (COVID-19 Data Repository Johns Hopkins University), and the number of infections per reported case in western Washington in this time frame was estimated to be 11.2 to 1. (Havers et al, 2020). Thus there were likely to have been ~13,000 COVID infections in King County by March 24.

161 of the cases through March 24 were sequenced, and 95% of were the D form at position 614 (GISAID data).

This was obviously a well-established D clade epidemic. Within two weeks, most of the infections sampled were G clade.

Fig 2. A plot of cumulative counts of D614 and G614 sequences by day in King county. Orange represents the original form, blue the form with the G614 mutation.

The declines in original-form frequencies and the increases in the G clade frequencies often continued well after stay-at-home orders were implemented, *g.* through April and into May. Again, this was repeated in many regions of the world, and it was nearly always the G form on the rise, not the D form. During the periods of regional lock down, founder effects through local super-spreader events and travel reintroductions would be minimized, and transmission within the community would play a more significant role (Korber et al. 2020, Figs. 2, S2, and S3).
An in-depth look at transmissibility in the UK, the most intensely sampled country globally, found that population genetic modelling indicated that 614G increases in frequency relative to 614D in a manner consistent with a selective advantage (Vogt et al. 2020). Using data prior to the lock down and samples that were derived only from clusters initiated in January or February, they found a selection coefficient of 0.21 (95% CI: 0.06 – 0.56). Using a separate fit in a period of epidemic decline, after April 15 and including clusters detected before March 31, they found a selection coefficient of 0.27 (95% CI:0.12–0.54). They did not find evidence for increased transmissibility of the 614G variant using phylodynamic analysis.
Finally, the simple fact that the original form was widely spread throughout the world by early March, but the G clade is now clearly the dominant form of the virus globally; and the transition took about 4–6 weeks.

Fig. 3. Maps showing the relative frequency of sampling G614 and G614 in different time windows. The size of the circle indicates the relative sampling in a given country within each of the four maps.

Higher Viral RNA levels in the upper respiratory tract in infection. The association between the G clade and higher levels of virus in the upper respiratory tract was measured by the number of RT PCR cycle thresholds required to detect the virus. It has now been reproduced in multiple labs, both our study, with close to 1000 patients sampled in Sheffield, England, and by two independent groups in the USA, in Chicago and in Washington state (Korber et al. 2020; Lorenzo-Redondo et al., 2020; Wagner et al., 2020). This measure has some dependence on the time from infection, and it is not a strict measure of infectious virus, rather of levels of viral

RNA, so it can be noisy. Still, that the G clade was significantly associated with higher levels of viral RNA reproducibly in multiple studies gives us confidence in the result.

G 614 Spike's greater infectivity in pseudotype virus assays. There is a clear change in phenotype between the original form and the G clade for the virus that has repeatedly been demonstrated in pseudotype assays. Data from two labs (Erica Ollmann-Saphire's lab and David Montefiori's lab) are included in Korber et al. 2020. Both laboratories used distinct assays and tested multiple cell types. At this point multiple pre-prints also support the greater infectivity of the G clade virus; the first study to show this was Zhang et al., 2020. This effect is highly significant, and reproducible, so we know these viruses are indeed different from each other by this measure.

What we still don't know is how this will translate into a controlled *in vivo* scenario; this is currently being explored by other groups.

4. What is the evidence that the G614 virus is not associated with greater disease severity?

Although G614 was associated with higher levels of viral RNA in the respiratory tract, we found no relationship between having this form of the virus and being hospitalized due to COVID-19. Others have also not seen such an association (e.g. Volz et al. 2020). A negative result does not prove that there is no relationship. It indicates that given the amount of data we had, and the way we were looking at it, we did not see an association. This is still good news because if there is an effect, it is likely to be subtle.

We do not know why G614 might be associated with higher levels of virus in the upper airway but not with a more severe pathology. It is possible that severe disease may be related to infection deeper in the respiratory tract, or that higher levels of virus results in more intense immune response that are better able to contain the infection.

Others have presented evidence that raises the possibility that G614 may be associated with a higher frequency of lethal outcomes using a very different approach, comparing deaths per known cases to G614 frequencies at a country level (Becerra-Flores and Cardozo, 2020). These results are interesting, but an important caveat regarding interpretation is that different countries used very different strategies to track cases and deaths.

5. How do you respond to criticism that the increases in the G clade frequency could have happened by chance alone?

The commentary on our *Cell* paper (Grubaugh 2020, *Cell* commentary on Korber et al. 2020) stated "*Over the period that G614 became the global majority variant, the number of introductions from China where D614 was still dominant were declining, while those from Europe climbed. This alone might explain the apparent success of G614.*"

Our findings were not just dependent on new introductions from China or Europe, and Europe was certainly not just one G clade epidemic, but a complex mixed epidemic in early March. There were many ongoing regional epidemics all over the world that were already well established by the Wuhan form in which community spread was ongoing, with many samples available. In almost all of these cases when the G clade first made its entrance and rose to high enough levels to be sampled, it was a matter of weeks for it to become the dominant form.

Below is the WHO situation report map of confirmed cases through March 15, and the frequency map for the two forms through the same period. There were ongoing epidemics in many places all over the world that were either predominantly the original form through mid-March, or evenly split with the original Wuhan form still very common. The Pandemic wasn't just being maintained by travelers who were moving out from China and Europe, but by community transmission networks and travelers from other places as well as China and Europe. Much of Europe was evenly split and some places even dominated by the original D614 form at that point in time (Spain and Wales, for example), as was almost all of Asia, and the Western United States. Within a month, it was almost gone.

Fig. 4. Top: WHO situation report showing confirmed cases by March 15, 2020, highlighting Europe. Bottom: cov.lanl.gov map of the relative frequencies of the G614 and D614 forms, also through March 15 and highlighting Europe.

Some people have suggested that the virus could have easily gotten lucky, and its spread be due to founder effects. We think the comprehensive epidemiological evidence across many regions makes this highly unlikely.

Obviously, a founder effect *could* have resulted in a change in frequency in a given population, or even a few populations. But to repeatedly drive out locally well-established epidemics via new introductions would require more than luck. It would require some sort of selective pressure to beat the odds and push 97% of the locations in only one direction. We systematically evaluated the changes in frequency in every single location where GISAID data showed both forms co-circulating and there were enough data and time to monitor a change: the G clade would have had to "get lucky" in almost every geographic region across the globe, wherever there was enough data to look; this is very unlikely to have happened by chance alone.

By the time our *Cell* paper was published, Fig. 1B and 3, available data included 50 geographic areas with significantly shifting frequencies, and among these were only 3 exceptions to the increase in G614. A statistical analysis of sets of independent regions showed that the probability is vanishingly small that such extreme repetition could be due to chance.

We highlighted these rare (D–form increasing) 3 exceptions in our *Cell* paper. The first exception was Iceland, which was readily explained by sampling biases (Korber et al., 2020, Fig. S4); no new samples have come in from Iceland since April. Here is an update on two others, Santa Clara County Public Health Department (Korber et al., 2020, Fig. S4) and Yakima (Korber et al., 2020, Fig. 3). Now both locations have been sampled for a longer time, and recent samples now both favor G form; they have gone from being exceptions to supporting our hypothesis. Here are the current versions of those plots (August 18, 2020). Both locations have now shifted to the G form.

Fig. 5. Two rare locations had an increase in the Wuhan form frequency identified in our comprehensive global scan of GISAID at the time of our *Cell* publication: Santa Clara County Public Health Department and Yakima Washington. More recently, with a more extended sampling time, they are also shifted towards the G form. This can be seen in the plots from August 18, 2020 which represent the weekly average running counts based on sequences in GISAID. Note the shift from orange (D) to blue (G) over time.

Furthermore, the regional frequency increases in the G clade steadily persisted well after regional stay at home orders were implemented and travel was restricted (Figure 2, Sup 2, and Sup 3, Korber et al. 2020).

Finally, Stephanie Pappas of Live Science (July 10, 2020) quoted Dr. Grubaugh as saying, "'*What's going to be important now is to continue to monitor in these places,'* Grubaugh said. If the G variant continues to dominate even in places where both the G and D versions are present, that might be a sign that the G mutation does provide the virus a transmission advantage."

What he suggests is "*going to be important now*" is in fact what we have been carefully and systematically documenting for the past months, what we set up the website cov.lanl.gov to do, and is what we published in the Cell paper.

In the controversy over the interpretation of the global shift to G614 viruses, some have argued that the mutation should not be studied. While we agree there are indeed many important issues to study, but we think this newer form of the virus is one of them. It is important both from an epidemiological point of view, and because of the new scientific insights and reagents it has already offered the field. (See question 6 below.)

There is room for many avenues of exploration of the biology of this virus.

6. Why do you think understanding the biology behind the greater transmissibility of the G form is important?

We did not originally know what the biological mechanism was that could give rise to a more transmissible virus. We hypothesized that it could have either been due to an immunological advantage (e.g. antibody resistance through allosteric effects or conformational change, or antibody enhancement as the 614 in embedded in an epitope that was implicated in enhancement for the first SARS CoV), or due to a fitness advantage in terms of infectiousness.

Given how rapidly this mutation was becoming the globally dominant form, it was important to understand why. The most urgent question was whether the mutation affected the antigenicity of the virus. This was critical because the D form of the virus was being used for most Spike-based vaccines, but the contemporary form of the virus that the vaccines needed to protect against was the G clade. It was important to establish whether the apparent increase in transmission was related to antibody resistance.

As it turns out it does. The G614 mutation significantly impacts antigenicity of the virus, but not at all in the way we were expecting. The G form of Spike is *more* sensitive to neutralizing antibodies than the D form. This was reproducible in mice, monkeys, and humans (Weissman 2020).

This can be understood through structural studies (Weissman 2020) and molecular dynamic studies (Mansbach 2020) that have demonstrated that the G form is preferentially found in a “one up” conformation, a conformation that reveals both the receptor binding domain (RBD) where ACE2 binds, and the epitopes for the potent neutralizing antibodies that bind to that region.

The greater accessibility of the RBD could also explain the greater infectivity that has now been confirmed in cell culture using pseudotype viruses in multiple laboratories (Korber et al. 2020, Zhang 2020).

Finally, a serendipitous bit of good fortune that came from studying G614 Spike is that its greater infectivity in a pseudotype assay enables a more robust assay with a stronger baseline signal, something the field was struggling with using the D614 form of Spike. So not only is G614 the more relevant form because it is the current form of the virus circulating in the world today (and therefore the form of the vaccine we need to protect against), it is also relevant because using G614 Spike has resulted in more a robust Spike pseudotype assay for comparing vaccine results as they unfold.

We now have several lines of epidemiological and experimental evidence that are all consistent with greater transmissibility of G614. Because we got the word out early, these experiments were started in April and

completed by early summer. We now have a better understanding of the virus, and a better understanding of the immunological implications of using D614 Spike vaccine in a G614 Spike pandemic. Also, the experimental reagents based on the currently prevalent and therefore relevant form of the virus are already in place in many laboratories; this is important because there indeed are phenotypic differences in the G614 form of the Spike that impact the assays.

Some of the open questions that are currently being explored by others include:

1. What is the impact of the G614 mutation and the other mutations in the G clade *in vivo* in animal models?
2. The G614 evaluated using pseudotype virus is more infectious in cell cultures. How will this increase translate in the context of the natural SARS-CoV-2 virus?
3. Since the G form spends more time in the "one-up" conformation (which exposes the receptor-binding domain and key neutralizing epitopes), would the G614 Spike make a better immunogen than the D614 form for inducing RBD antibodies?
4. Does the higher infectivity and greater neutralization sensitivity of the G clade have additional unforeseen *in vivo* effects regarding vaccine protection? Animal vaccine studies using either D or G vaccines should be tested with G clade challenge viruses; the first of these studies used only D vaccines and D challenge viruses.
5. Do any of the other mutations that define the G clade have phenotypic consequences? The mutations in RdRp and the 5' UTR might.

We had two motivations for submitting a preprint regarding the G614 mutation. The first was to alert people in the immunology/virology community, so they could start to work on it if they were interested. The second motivation was to let people know about the accumulating evidence indicating that the G clade, the virus becoming prevalent in the late spring and early summer, was likely to be more transmissible than the original Wuhan form that dominated much of the global epidemic in the early spring.

7. Can you use frequencies to look for signs of evolutionary positive selection?

Yes; this strategy has an old and venerable tradition, fully fleshed out over decades (see Endler, 1986). Frequencies of distinct phenotypes are a mainstay of evolutionary thought: think back to the peppered moth that many of us learned about in high school biology. The light-colored moths were able to blend in naturally with light-colored tree bark and lichen, so the moths were camouflaged from the birds that preyed on them. But

during the industrial revolution, in regions where tree bark was covered with soot, a black phenotype was better camouflaged and became the dominant variant. The science behind this was ultimately validated in an extensive field study conducted by Michael Majerus; his work was published posthumously (Cook et al. 2012). Here are a few additional examples of using frequency shifts to monitor natural selection (Linnen and Hoekstra 2009 and Barrett et al. 2019).

Given the data available from GISAID (sequence, geographic and sampling data), we considered a frequency-based analysis to be a reasonable strategy, providing that we attended to the possibility of founder effects and sought independent experimental confirmation. We did this carefully. We required that frequency shifts were towards the G clade in almost all geographic areas with enough data to look, and we had found many examples with strikingly repetitive pattern; that the shift occurred even in the context of very well-established D form local epidemics: the shift continued well after stay-at-home orders were imposed (point 3 above for details). By the time we had had published in the bioRxiv, we had clinical data from our colleagues in Sheffield England showing that viral RNA levels in the upper respiratory tract were associated with the G clade in clinical samples, a statistically supported phenotypic distinction that was consistent with the possibility of increased transmissibility. By the time we published in *Cell*, our colleagues at Duke and the La Jolla Institute of Immunology had also shown that the G form of Spike was more infectious.

We did not use phylogenetic homoplasy-based methods to look for evidence of positive selection: the G614 mutation is transmitted from person to person as part of the G clade, with a shared common ancestor, >99.99% of the time — it seldom arises from a *de novo* mutation. In this situation, homoplasy-based methods are inappropriate, as they rely on repeated mutations. Although this class of methods is not a useful strategy for exploring the hypothesis that the G614 mutation might be under positive selection, such methods may be helpful for identifying positive selection in other contexts among COVID-19 pandemic genetic variants as the pandemic progresses. Furthermore, we were concerned about the potential for recombination, and think it should be explicitly addressed in any such attempts in future studies.

8. What is a pseudovirus?

Pseudotype virus neutralization measurements are highly correlated with authentic SARS-CoV-2 measurements (Schmidt et al. 2020). Pseudoviruses are mimics of the live virus that are engineered to be capable of only a single round of infection as a safety measure for laboratory workers. They provide a safe, quantitative, and reproducible way to work with and directly compare proteins that mediate viral entry into cells. For SARS-CoV-2, that would be the Spike protein, while for HIV-1 that would be the

Envelope protein. These pseudovirus assay systems are essential for assessing vaccine responses and determining which antibodies might be best for antibody therapeutics, and so they were being developed for SARS-CoV-2 immune response testing over the course of the spring in many laboratories.

Pseudoviruses are made from various different laboratory viruses that are hobbled for safety. Their native entry proteins are replaced by those of the virus under study. Our experiments used two different viral backbones, with SARS-CoV-2 Spike inserted; they can infect human cells via Spikes interactions with ACE2, but, by design, cannot reproduce (and are hence safer and easier to work with in the laboratory). A pseudovirus can carry a marker to enable researchers to readily count how many cells in a culture are infected by a certain amount of virus. This enables scientists to both compare the infectivity of different Spikes and to quantitatively determine how well different antibodies block viral entry into cells under highly standardized conditions.

These assays are now in place and ready to go, but there were many challenges in getting the experiments going for the new COVID-19 virus. Drs. Ollmann Saphire and the Montefiori labs (both are co-authors on our *Cell* paper) lead central testing labs where they help explore the effectiveness of different immune interventions to block the virus and prevent infection; this was their primary motivation in getting the pseudovirus systems working smoothly in their labs. Their work over the coming year will help resolve which antibodies are best for therapeutics and which vaccines hold the most promise. They both were interested in testing the original Wuhan form of the virus as well as the G614 variant, and using their respective assay systems found significant differences with the G614 variant being more infectious than the Wuhan form.

Testing with live viruses is an important counterpoint to pseudoviruses. Studies to compare G614 and D614 forms in the context of natural SARS-CoV-2 viruses are underway by others.

9. Can a single amino acid alter the phenotype of a protein?

Yes. There are vast numbers of clear examples of this in the scientific literature. Sickle-cell anemia may be the most well-known example among the general population.

In particular, when a virus infects a new host, it may be subject to adaptations that enable it to propagate better in that host. SARS-CoV-2 mutations are rare, but there have been over 18 million COVID-19 cases in the world, so there is ample opportunity for an advantageous mutation to arise.

Here are just three examples of one change making a difference that are relevant to the study at hand:

Immune escape, SARS-CoV-1 example:

Broadening of Neutralization Activity to Directly Block a Dominant Antibody-Driven SARS-Coronavirus Evolution Pathway. Sui et al. PLoS Pathog. (2008) 4: e1000197:

In this case a SARS-Cov Spike RDB mutation naturally arose between 2002/3 and 2003/4 in civet cats. It conferred resistance to the potent RDB targeting NAb 80R. The escape mutation evolution could be recapitulated in vitro. Awareness of the mutation and its impact enabled the authors to devise ways to mitigate its impact.

Enhanced infectivity, an HIV-1 example:

A signature in HIV-1 envelope leader peptide associated with transition from acute to chronic infection impacts envelope processing and infectivity. Asmal et al. PloS One (2011) 6:e23673.

Host specificity, HIV-1 example:

Envelope residue 375 substitutions in simian-human immunodeficiency viruses enhance CD4 binding and replication in rhesus macaques. Li et al. PNAS USA (2016) 113:E3413

10. Does the G clade represent a new viral strain?

In our original preprint we used the word "strain" to refer to viruses that carried the G614D; some people objected to this usage (see: Ed Yong, May 6, 2020).

Wikipedia gives a loose definition — "In biology, a strain is a genetic variant, a subtype or a culture within a biological species." This usage can be readily found in the scientific literature, where "strain" is often used interchangeably with "variant." Here are a few of many examples: Lee, JM. et al. 2019, Krakoff et al. 2019, and Worobey et al. 2020.

GISAID's website uses this more general definition of strain. Here are two COVID-19 examples: "*Depending on choice of definitions one can classify the circulating virus strains into a different number of clades based on genetic variants*." (In: Natural evolution of the hCoV-19 virus, April 13, 2020, GISAID), and "*For each strain the clade information is provided in the "Virus detail" section of the metadata."* (In: Clade and lineage nomenclature, July 4, 2020, GISAID*).*

But many feel strongly that the word strain should be used only if the genetic variant is associated with unique phenotypic characteristics (e.g. Kuhn et al., Arch Virol. 158:301). We would argue that the G clade also

meets this second stricter designation, as the phenotype of the virus is significantly and reproducibly distinct from the Wuhan reference virus, but that this is not important.

The important point to us is not the use of the word "strain", but the biology. The phenotype of the G clade virus is in fact clearly distinguishable, and this has now been shown at many levels and reproduced in many different laboratories. Although we used the term "strain" in our *bioRxiv* preprint, we did not use it in our *Cell* paper, hoping to refocus the broader discussion away from semantics and towards the data and its implications.

Meanwhile, since our Cell paper, the evidence for a phenotypic difference in the G clade virus relative to the D clade has grown ever stronger:

1. The statistical support in the epidemiological data grew stronger every week through the spring, as more and more geographic areas went from D form prevalent to G form prevalent.
2. The Ct measure showing the association of G614 with higher levels of viral RNA in patients has been repeated independently in multiple laboratories.
3. The G614 variant has now been shown to be more infectious in multiple pseudotyping assays in multiple laboratories.
4. The G614 mutation has also been shown to be highly significantly associated with *greater* sensitivity to neutralizing antibodies, both vaccine-induced and in sera from natural infection. This was a phenotypic difference we didn't anticipate, but were delighted to see! (Weissman 2020).
5. Spikes with the G614 mutation have been shown both structurally (Weissman 2020) and through molecular dynamics (Gnanakaran 2020) to prefer a one-up conformation which increases the ACE2 receptor binding site accessibility as well as enhances exposure of RBD antibody epitopes. These structural studies illuminate a single mechanism that could explain both the increased infectivity and the enhanced sensitivity to neutralizing antibodies of the G clade.

11. Can the word "mutation" be applied to an amino acid change?

Mutations arise when genetic material (DNA or RNA) is miscopied; in the case of corona viruses, RNA is the genetic material. A base change that results in a different amino acid being incorporated into a protein is called a missense or nonsynonymous mutation.

Some people prefer to use the word mutation only to describe changes at the DNA or RNA level. The *mutation* from A-to-G at position 23,403 gives rise to the amino acid change in the Spike protein at position 614 from D-to-G.

But we, *and many other scientists and scientific journals*, do not subscribe to this usage restriction. For example, a quick search of PubMed for the terms “amino acid mutation” or “amino acid mutations” yields 1,735 scientific papers. A Google Books ngram search shows comparable usage for the terms “nucleotide mutation” and “amino acid mutation.”

Fig. 6. Usage of the term amino acid mutation vs nucleotide mutation.

We term “amino acid mutation” in this Q & A, however, and will continue to do so in other settings, as we find this to be concise and to accurately convey our meaning.

References

Becerra-Flores, M., and Cardozo, T. (2020). *SARS-CoV-2 viral spike G614 mutation exhibits higher case fatality rate.* Int J Clin Pract, e13525.

Barrett et al. 2019. *Linking a mutation to survival in wild mice* Science 2019 363: 499-504

Cook et al. 2012. Cook, L. M.; Grant, B. S.; Saccheri, I. J.; Mallet, James (2012). *Selective bird predation on the peppered moth: the last experiment of Michael Majerus. Biology Letters.* 8 (4): 609–612. doi:10.1098/rsbl.2011.1136. PMC 3391436. PMID 22319093

Endler, JA. (1986) Natural Selection in the Wild. Volume 21, Monographs in Population Biology, Princeton University Press.

Havers et al. 2020*. Seroprevalence of Antibodies to SARS-CoV-2 in 10 Sites in the United States, March 23-May 12, 2020.* Havers et al. JAMA July online ahead of print July 21.

Korber et al., 2020. *Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell* 182, 1-16 (2020).

Korber et al., 2020, BioRxiv. *Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2.* bioRxiv April 30 doi: https://doi.org/10.1101/2020.04.29.069054

Krakoff E, Gagne RB, VandeWoude S, Carver S. 2019 *Variation in Intra-individual Lentiviral Evolution Rates: a Systematic Review of Human, Nonhuman Primate, and Felid Species.*J Virol . 2019 Jul 30;93(16):e00538-19. doi: 10.1128/JVI.00538-19. Print 2019 Aug 15.

Lee, JM. et al. 2019. *Mapping person-to-person variation in viral mutations that escape polyclonal serum targeting influenza hemagglutinin* eLife. 2019; 8: e49324. PMID: 31452511

Linnen and Hoekstra 2009. *Measuring Natural Selection on Genotypes and Phenotypes in the Wild.* Cold Spring Harb Symp Quant Biol 2009. 74: 155-

168

Lorenzo-Redondo, R., Nam, H.H., Roberts, S.C., Simons, L.M., Jennings, L.J., Qi, C., Achenbach, C.J., Hauser, A.R., Ison, M.G., Hultquist, J.F., et al. (2020). A Unique Clade of SARS-CoV-2 Viruses is Associated with Lower Viral Loads in Patient Upper Airways. medRxiv, 2020.2005.2019.20107144.

Mansbach et al. 2020. *The SARS-CoV-2 Spike Variant D614G Favors an Open Conformational State,* bioRxiv 2020 (**doi**: https://doi.org/10.1101/2020.07.26.219741)

Pappas, 2020. *A new coronavirus mutation is taking over the world. Here's what that means.* By Stephanie Pappas - Live Science Contributor. https://www.livescience.com/new-coronavirus-mutation-explained.html

Schmidt et al. J Exp Med 2020 Nov 2;217(11):e20201181. Measuring SARS-CoV-2 neutralizing antibody activity using pseudotyped and chimeric viruses

Voltz et al. 2020. *Evaluating the effects of SARS-CoV-2 Spike mutation D614G on transmissibility and pathogenicity.* Erik M Volz, Verity Hill, John T McCrone, Anna Price, David Jorgensen, Aine O'Toole, Joel Alexander Southgate, Robert Johnson, Ben Jackson, Fabricia F. Nascimento, Sara M. Rey, Samuel M. Nicholls, Rachel M. Colquhoun, Ana da Silva Filipe, Nicole Pacchiarini, Matthew Bull, Lily Geidelberg, Igor Siveroni, Ian G. Goodfellow, Nicholas James Loman, Oliver Pybus, David L Robertson, Emma C Thomson, Andrew Rambaut, Thomas R Connor, The COVID-19 Genomics UK Consortium. medRxiv 2020.07.31.20166082; doi: https://doi.org/10.1101/2020.07.31.20166082

Wagner, C., Roychoudhury, P., Hadfield, J., Hodcroft, E., Lee, J., Moncla, L., Muller, N., Behrens, C., Huang, M.-L., Mathias, P., et al. (2020). Comparing viral load and clinical outcomes in Washington State across D614G mutation in spike protein of SARS-CoV-2.

Weissman D. et al. 2020. D614G *Spike Mutation Increases SARS CoV-2 Susceptibility to Neutralization.* medRxiv (doi: https://doi.org/10.1101/2020.07.22.20159905)

Worobey et al. 2020. Worobey, M. Plotkin S., and Hensley S. *Influenza Vaccines Delivered in Early Childhood Could Turn Antigenic Sin into Antigenic Blessings.* Cold Spring Harbor Perspect Med 2020 Jan 21;a038471

Zhang et al. 2020. Zhang L, Jackson CB, Mou H, Ojha A, Rangarajan ES, Izard T, Farzan M, Choe H. The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. bioRxiv 2020.06.12.148726; doi: https://doi.org/10.1101/2020.06.12.148726

Newer variant of COVID-19-causing virus dominates global infections

Virus with D614G change in Spike out-competes original strain, but may not make patients sicker

LOS ALAMOS, N.M., July 2, 2020— Research out today in the journal Cell shows that a specific change in the SARS-CoV-2 coronavirus virus genome, previously associated with increased viral transmission and the spread of COVID-19, is more infectious in cell culture. The variant in question, D614G, makes a small but effective change in the virus's 'Spike' protein, which the virus uses to enter human cells.

Bette Korber, a theoretical biologist at Los Alamos National Laboratory and lead author of the study, noted, "The D614G variant first came to our attention in early April, as we had observed a strikingly repetitive pattern. All over the world, even when local epidemics had many cases of the original form circulating, soon after the D614G variant was introduced into a region it became the prevalent form."

Geographic information from samples from the GISAID COVID-19 viral sequence database enabled tracking of this highly recurrent pattern, a shift in the viral population from the original form to the D614G variant. This occurred at every geographic level: country, subcountry, county, and city.

Two independent lines of experimental evidence that support these initial results are included in today's paper. These additional experiments, led by Professor Erica Ollmann Saphire, Ph.D., at the La Jolla Institute, and by Professor David Montefiori, Ph.D., at Duke University, showed that the D614G change increases the virus's infectivity in the laboratory. These new experiments, as well as more extensive sequence and clinical data and improved statistical models, are presented in the Cell paper. More in vivo work remains to be done to determine the full implications of the change.

The SARS-CoV-2 virus has a low mutation rate overall (much lower than the viruses that cause influenza and HIV-AIDS). The D614G variant appears as part of a set of four linked mutations that appear to have arisen once and then moved together around the world as a consistent set of variations.

"It's remarkable to me," commented Will Fischer of Los Alamos, an author

on the study, “both that this increase in infectivity was detected by careful observation of sequence data alone, and that our experimental colleagues could confirm it with live virus in such a short time.”

Fortunately, “the clinical data in this paper from Sheffield showed that even though patients with the new G virus carried more copies of the virus than patients infected with D, there wasn’t a corresponding increase in the severity of illness," said Saphire, who leads the Gates Foundation-supported Coronavirus Immunotherapy Consortium (CoVIC).

Korber noted, “These findings suggest that the newer form of the virus may be even more readily transmitted than the original form – whether or not that conclusion is ultimately confirmed, it highlights the value of what were already good ideas: to wear masks and to maintain social distancing.”

Research partners from Los Alamos National Laboratory, Duke University, and the University of Sheffield initially published work on this analysis on the bioRxiv site in an April 2020 preprint. That work also included observations of COVID-19 patients from Sheffield that suggested an association of the D614G variant with higher viral loads in the upper respiratory tract.

“It is possible to track SARS-CoV-2 evolution globally because researchers worldwide are rapidly making their viral sequence data available through the GISAID viral sequence database”, Korber said. Currently tens of thousands of sequences are available through this project, and this enabled Korber and the research team to identify the emergence of the D614G variant.

GISAID was established to encourage collaboration among influenza researchers, but early in the epidemic the consortium established a SARS-CoV-2 database, which soon became the de facto standard for sharing outbreak sequences among researchers worldwide.

The study, "Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus" (DOI: 10.1016/j.cell.2020.06.043) was supported by the Medical Research Council (MRC) part of UK Research & Innovation (UKRI the National Institute of Health Research (NIHR); Genome Research Limited, operating as the Wellcome Sanger Institute; CoVIC, INV-006133 of the COVID-19 Therapeutics Accelerator, supported by the Bill and Melinda Gates Foundation, Mastercard, Wellcome; private philanthropic support, as well as the Overton family; a FastGrant, from Emergent Ventures, in aid of COVID-19 research; and the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Interagency Agreement No. AAI12007-001-00000, and the Los Alamos Laboratory Directed Research and Development program.

Additional study authors included S. Gnanakaran, H. Yoon, J. Theiler, W. Abfalterer, N. Hengartner, E.E. Giorgi, T. Bhattacharya, B. Foley, K.M. Hastie, M.D. Parker, D.G. Partridge, C.M. Evans, T.M. Freeman, T.I. de Silva, C. McDanal, L.G. Perez, H. Tang, A. Moon-Walker, S.P. Whelan, C.C. LaBranche.

About Los Alamos National Laboratory
Los Alamos National Laboratory, a multidisciplinary research institution engaged in strategic science on behalf of national security, is managed by Triad, a public service oriented, national security science organization equally owned by its three founding members: Battelle Memorial Institute (Battelle), the Texas A&M University System (TAMUS), and the Regents of the University of California (UC) for the Department of Energy's National Nuclear Security Administration.

Los Alamos enhances national security by ensuring the safety and reliability of the U.S. nuclear stockpile, developing technologies to reduce threats from weapons of mass destruction, and solving problems related to energy, environment, infrastructure, health, and global security concerns.

LA-UR-20-26388

Video

Abstract

Tracking SARS-CoV-2 Spike mutations: evidence for increased infectivity of D614G

Summary. A SARS-CoV-2 variant carrying the Spike protein amino acid change D614G has become the most prevalent form in the global pandemic. Dynamic tracking of variant frequencies revealed a recurrent pattern of G614 increase at multiple geographic levels: national, regional and municipal. The shift occurred even in local epidemics where the original D614 form was well established prior to the introduction of the G614 variant. The consistency of this pattern was highly statistically significant, suggesting that the G614 variant may have a fitness advantage. We found that the G614 variant grows to higher titer as pseudotyped virions. In infected individuals G614 is associated with lower RT- PCR cycle thresholds, suggestive of higher upper respiratory tract viral loads, although not with increased disease severity. These findings illuminate changes important for a mechanistic understanding of the virus, and support continuing surveillance of Spike mutations to aid in the

development of immunological interventions.

Media

Did a Mutation Help the Coronavirus Spread? More Evidence, but Lingering Questions —**The New York Times** (7/2)
Researchers claim that a predominating variant had a “fitness advantage.” But many experts are not persuaded.

New form of coronavirus spreads feaster, but doesn't make people sicker —**CNN** (7/2)
A global study has found clear evidence that a new form of the coronavirus has spread from Europe to the US. The new mutation makes the virus more infectious but does not seem to make people any sicker, an international team of researchers reported.

This coronavirus mutation has taken over the world. Scientists are trying to understand why. —**Washington Post** (7/2)
At least five laboratory experiments suggest that the mutation makes the virus more infectious, although only one of those studies has been peer-reviewed. That study, led by scientists at Los Alamos National Laboratory and published Thursday in the journal Cell, also asserts that patients with the G variant actually have more virus in their bodies, making them more likely to spread it to others.

The coronavirus has changed since it left Wuhan. Is it more infectious? — **Los Angeles Times** (7/2)
The study authors, led by Bette Korber, a computational biologist at Los Alamos National Laboratory, posted a preliminary version of the work in May that generated substantial controversy by claiming the mutation in the spike protein made the virus more contagious.

Newer, More Dominant COVID-19 Variant Is More Infectious in the Lab — **Genetic Engineering & Biotechnology News** (7/2)
“The D614G variant first came to our attention in early April, as we had observed a strikingly repetitive pattern," said Bette Korber, the study’s lead author and a theoretical biologist at Los Alamos National Laboratory.

Social Media

twitter @LosAlamosNatLab

Tweets by LosAlamosNatLab

facebook @LosAlamosNationalLab

Los Alamos National Laboratory